Focus Article

Subspace clustering

Hans-Peter Kriegel,* Peer Kröger and Arthur Zimek

Subspace clustering refers to the task of identifying clusters of similar objects or data records (vectors) where the similarity is defined with respect to a subset of the attributes (i.e., a subspace of the data space). The subspace is not necessarily (and actually is usually not) the same for different clusters within one clustering solution. In this article, the problems motivating subspace clustering are sketched, different definitions and usages of subspaces for clustering are described, and exemplary algorithmic solutions are discussed. Finally, we sketch current research directions.

How to cite this article:
WIREs Data Mining Knowl Discov 2012, 2: 351–364 doi: 10.1002/widm.1057

INTRODUCTION

Clustering is the problem of finding a set of groups of similar objects within a data set while keeping dissimilar objects separated in different groups or the group of noise. With the development of modern capabilities of automatic data generation and acquisition in more and more application domains, a huge amount of data is produced—huge not only in terms of the number of data objects but also in terms of the number of attributes of each object. High-dimensional data, however, pose new challenges to the clustering problem that requires specialized solutions which have only begun to be addressed since late 1990s. Two major issues can be distinguished in the clustering problem in general, namely (1) the adopted paradigm and algorithmic approach to clustering and (2) the definition and assessment of similarity versus dissimilarity. The former is discussed in other articles concerned with clustering paradigms. The latter is the key issue in subspace clustering. Different weighting, different selections, or different combinations of attributes of a data set are equivalent to defining or deriving different subspaces to express appropriately the desired properties of a model of similarity suitable to a given application domain. For different clusters within one and the same clustering solution, usually different subspaces are relevant. Thus subspace clustering algorithms cannot be thought of as variations of usual clustering algorithms using just a different definition of similarity. Instead, the similarity measure and the clustering solution are derived simultaneously and depend on each other.

Surveys on the topic of subspace clustering are Refs 1–4. Parsons et al.[1] survey the problem rather illustratively. In Ref 2, a concise theoretical overview is provided. In Ref 3, the problem is discussed in depth, along with different subproblems and a number of example algorithms. In the recent article,[4] the problem is surveyed from the point of view of machine learning and of computer vision. Some recent textbooks[5,6] already sketch some example algorithms and touch upon some problems. Recent experimental evaluation studies covered some selections of a specific subtype of subspace clustering, algorithms for clustering in axis-parallel subspaces.[7,8] In this focus article, we describe different kinds of subspaces that have been reflected in different approaches to subspace clustering. We will discuss dependencies and differences between these kinds of subspaces and the corresponding subspace clustering algorithms. For each category, we will discuss only some exemplary approaches. Note that the term 'subspace clustering', in a narrower sense, does also relate to a special category of clustering algorithms in axis-parallel subspaces (which will be discussed below) but we study the wider meaning of this term here as relating to different kinds of clustering algorithms in different kinds of subspaces.

ASPECTS OF THE PROBLEM

The 'Curse of Dimensionality'

Designing specialized solutions for analyzing high-dimensional data has often been motivated by the so-called *curse of dimensionality*. This term, however,

*Correspondence to: kriegel@dbs.ifi.lmu.de

Institut für Informatik, Ludwig-Maximilians-Universität München, Oettingenstrasse 67, D-80538 M, Munich, Germany

DOI: 10.1002/widm.1057

relates to a bundle of rather different problems. Some of these are elaborated in more detail in the literature concerned with index structures.[9] Here, we highlight aspects of the problem that are often seen in relation to subspace clustering although they are of different nature and importance.[3]

Aspect 1: Optimization Problem

Historically, the first aspect has been described in Ref 10 in the context of optimization problems. Clearly, the difficulty of any global optimization approach increases exponentially with an increasing number of variables (dimensions). This problem is particularly well known in the area of pattern recognition. At a general level, this problem relates to the clustering problem. Seeking a clustering of a data set supposes the data being generated by several functions. Ideally, a clustering model would enable the user to identify the functional dependencies resulting in the data set at hand and, thus, to eventually find new and interesting insights in the laws of the domain the data set describes. Those functions are the more complex the more attributes contribute to the actual relationships.

Aspect 2: Concentration Effect of L_p Norms

The second aspect of the curse of dimensionality is the *deterioration of expressiveness* of the most commonly used similarity measures, the L_p norms, with increasing dimensionality. In their general form, for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, L_p norms are given as

$$\|\boldsymbol{x} - \boldsymbol{y}\|_p = \sqrt[p]{\sum_{i=1}^{d} |x_i - y_i|^p}. \tag{1}$$

The choice of p is crucial in high-dimensional data according to several studies.[11–13] The key result of Ref 11 states that if the ratio of the variance of the length of any point vector $\boldsymbol{x} \in \mathbb{R}^d$ (denoted by $\|\boldsymbol{x}\|$) to the length of the mean point vector (denoted by $E[\|\boldsymbol{x}\|]$) converges to zero with increasing data dimensionality, then the proportional difference between the farthest point distance $D_{\max}$ and the closest point distance $D_{\min}$ (the *relative contrast*) vanishes:

$$\lim_{d\to\infty} \operatorname{var}\left(\frac{\|\boldsymbol{x}\|}{E[\|\boldsymbol{x}\|]}\right) = 0 \Longrightarrow \frac{D_{\max} - D_{\min}}{D_{\min}} \rightarrow_p 0. \tag{2}$$

The precondition has been described as being valid for a broad range of data distributions. Intuitively, the relative contrast of (L_p) distances does diminish as the dimensionality increases. This *concentration effect* of the distance measure reduces the utility of the measure for discrimination.

This aspect of the problem is often quoted in motivating subspace clustering as a specialized procedure. It should be noted, though, that the problem is neither well enough understood (see, e.g., Ref 14) nor actually relevant when the data follow different, well separated distributions,[15–17] because in that case the precondition does not hold. Regarding the separation of clusters, the following aspects are far more important in the scenario of subspace clustering.

Aspect 3: Irrelevant Attributes

A third aspect is often confused with the previous aspect but is actually independent. To find dependencies and laws describing some occurring phenomena a glut of data is collected and single entities are described with many possibly related attributes. Among the features of a high-dimensional data set, for any given query object, many attributes can be expected to be irrelevant to describing that object. Irrelevant attributes can interfere with the performance of similarity queries for that object. The relevance of certain attributes may differ for different groups of objects within the same data set. Thus, because groups of data are defined by some of the available attributes only, many irrelevant attributes may interfere with the efforts to find these groups. Irrelevant attributes are often also referred to as 'noise'.

Aspect 4: Correlated Attributes

Similarly, as a fourth aspect, in a data set containing many attributes, there may be some correlations among subsets of attributes. In sight of feature reduction methods, all but one of these attributes may be redundant. However, from the point of view of a domain scientist who collected these attributes in the first place, it may be an interesting new insight that there are so far unknown connections between features.[18] As for the previous problem, this phenomenon may be differently relevant for different subgroups of a data set as correlations among attributes that are characteristic for a cluster will be different for other clusters of the data set.

In view of spatial queries, the observation that the intrinsic dimensionality of a data set usually is lower than the embedding dimensionality (based on interdependencies among attributes) has often been attributed to overcome the 'curse of dimensionality' (see, e.g., Ref 19). In view of the clustering problem, however, the aforementioned aspects remain unaffected by this phenomenon.

Consequences

The third and fourth of these aspects of the curse of dimensionality have been the main motivation for developing specialized methods for clustering in *subspaces* of potentially high-dimensional data. Since decades, research in statistics was concerned with the effect of irrelevant ('noise') attributes (sometimes called 'masking variables'[20]) on the performance of clustering algorithms.[21–24] They discuss how to identify such noise attributes that do not contribute to the cluster structure at all. The clustering task should then concentrate on the remaining attributes only. In some cases, the contribution of attributes may be weighted instead of a discrete decision (contributing or masking). This view of the problem is obviously closely related to dimensionality reduction.[25,26]

The field of subspace clustering as we sketch it here assumes a more complex view. The reasoning is that dimensionality reduction as a first step prior to clustering is not always resolving these issues because the correlation of attributes or the relevance versus irrelevance of attributes is usually characteristic for some clusters but not for complete data sets. In other words, 'masking variables' may be masking certain clusters but may help to discover others (see a simple example in Figure 1—feature weighting and selection may have stopped here, whereas subspace clustering would inspect the three-dimensional space for more discriminative subspaces).

The phenomenon that different features or a different correlation of features may be relevant for varying clusters has been called *local feature relevance* or *local feature correlation*.[3] Feature selection or dimensionality reduction techniques are *global* in the following sense: they generally compute only one subspace of the original data space in which the clustering can then be performed. In contrast, the problem of *local feature relevance* and *local feature correlation* describes the observation that multiple subspaces are needed because each cluster may exist in a different subspace. The critical requirement for the design of subspace clustering algorithms is hence the integration of some heuristic for deriving the characteristic subspace for a certain subgroup of a data set with some heuristic for finding a certain subgroup of a data set that exhibits a sufficient level of similarity in a certain subspace of the data space—obviously a circular dependency of two subtasks (subspace determination versus cluster assignment) already exist in the most basic problem description.

TYPES OF SUBSPACES

Although one has to keep in mind all the aspects of the curse of dimensionality sketched above, mainly motivating specialized approaches to clustering in subspaces are the phenomena of irrelevant attributes (aspect 3) and correlated attributes (aspect 4). These two phenomena result in different types of subspaces. It should be mentioned that the type of the considered subspaces depends on the employed clustering method.

Axis-Parallel Subspaces

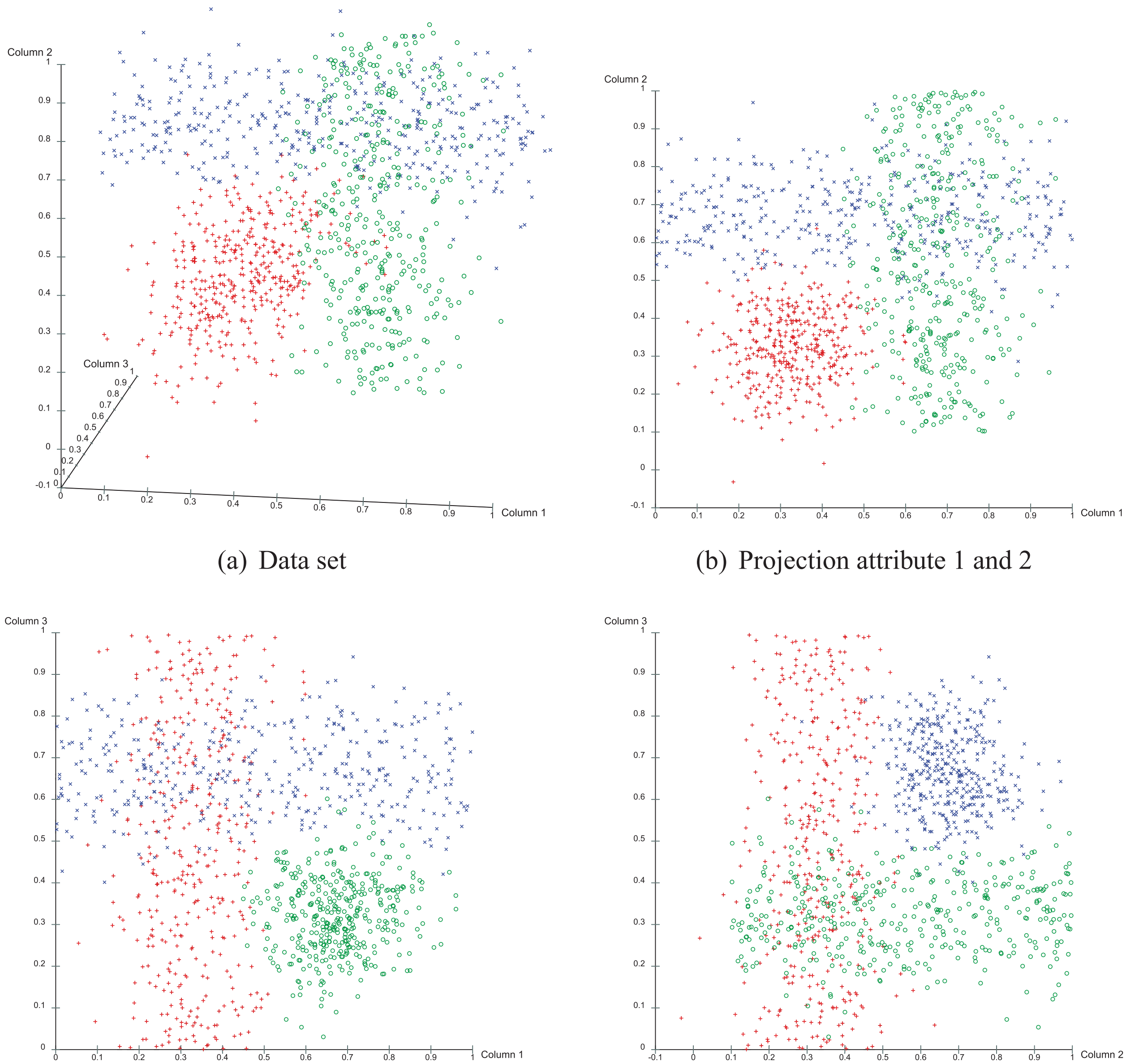
The distinction between relevant and irrelevant attributes generally assumes that the variance of attribute values for a relevant attribute over all points of the corresponding cluster is rather small compared to the overall range of attribute values, whereas the variance for irrelevant attributes within a given cluster is high or indistinguishable from the values of the same attribute for other clusters. For example, one could assume a relevant attribute for a given cluster being normally distributed with a small standard deviation, whereas irrelevant attribute values are uniformly distributed over the complete data space. The geometrical intuition of these assumptions relates to the points of a cluster being widely scattered in the direction of irrelevant axes while being densely clustered along relevant attributes. When selecting the relevant attributes only, the cluster would appear as a full-dimensional cluster in this subspace. In the full-dimensional space (including also the irrelevant attributes) the cluster points form a hyperplane parallel to the irrelevant axes. Because of this geometrical appearance, this type of cluster is addressed as 'axis-parallel subspace cluster' (see Figure 2).

Arbitrarily Oriented Subspaces

If two attributes a_i and a_j are correlated for a set of points, the points will be scattered along a hyperplane defined by some linear dependency between both attributes that corresponds to the correlation. The subspace orthogonal to this hyperplane is then a subspace where the points cluster densely irrespective of the variance of combined values of a_i and a_j. This subspace is arbitrarily oriented (see Figure 3) and, hence, the more general case compared to axis-parallel subspaces.

Special Cases

Although the aforementioned types of subspaces have a direct relationship to aspects of the curse of dimensionality as discussed above, it should be noted that from the view of existing clustering algorithms there



(a) Data set (b) Projection attribute 1 and 2

(c) Projection attribute 1 and 3 (d) Projection attribute 2 and 3

FIGURE 1 | Local feature relevance: three clusters, each cluster defined in two attributes, yet no any two clusters share two relevant attributes. Figures created with ELKI 0.4.[27]

are special types of subspaces. In fact, there is a large family of algorithms, referred to as 'biclustering', 'co-clustering', 'two-mode clustering', or 'pattern-based clustering'. Considering the spatial intuition of subspaces for subspace clustering, these algorithms address different kinds of subspaces. Thus, the types of the subspaces considered by these methods do not directly relate to the phenomena of the curse of dimensionality discussed here but rather are the products of specific cluster models. Refer to Ref 28 for an overview on biclustering methods. The relationship or correspondence between cluster models and types of subspaces resulting (implicitly) from the cluster model has been discussed in Ref 3.

CLUSTERING IN AXIS-PARALLEL SUBSPACES

Cluster Model

So far, most research in this field mainly transferred the full-dimensional cluster models of different

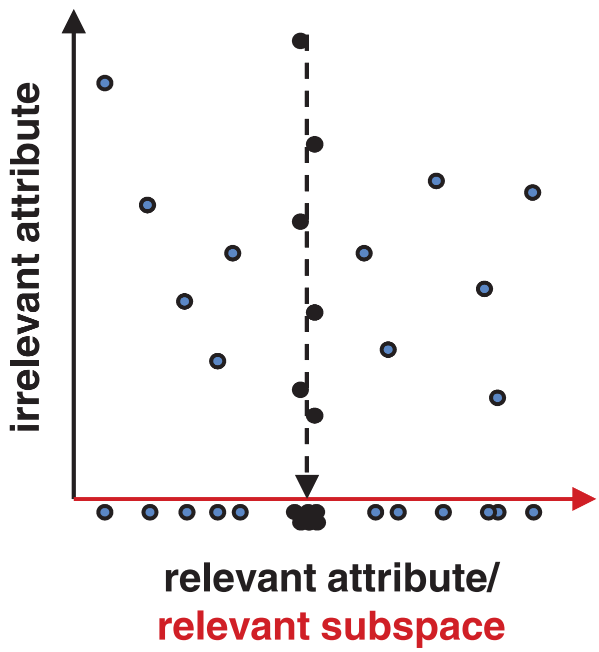


FIGURE 2 | Axis-parallel subspace cluster.

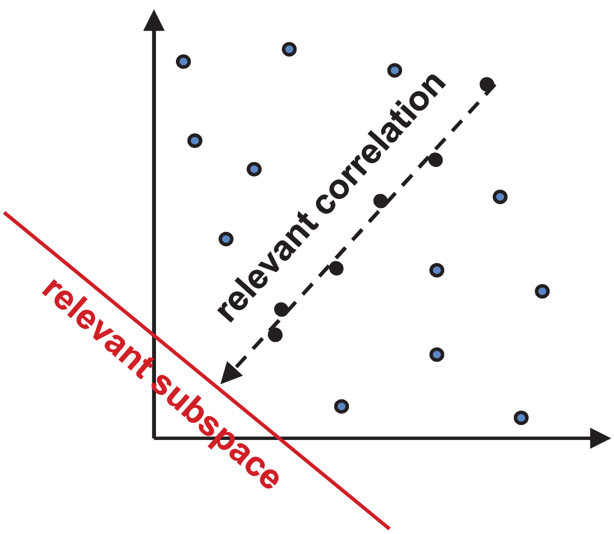


FIGURE 3 | Arbitrarily oriented subspace cluster.

clustering techniques into some subsets of, more or less, interesting subspaces of a data space. Hence there is no established 'subspace cluster model' describing axis-parallel subspace clusters concisely. Instead, there are only cluster models describing clusters in a subspace as if the corresponding subspace were the full-dimensional data space. The acclaimed goal of the proposed algorithms has apparently been defined in accordance with the algorithmic technique, that is, the problem has been defined according to the proposed solution. This methodical flaw has still repercussions for current research. The first statistically sound model for axis-parallel subspace clusters was proposed in Ref 29. It is based on the assumption that values of a relevant attribute for a subspace cluster are *not* uniformly distributed over the complete attribute domain and that this property satisfies a certain statistical test.

Basic Techniques

The number of possible axis-parallel subspaces where clusters could reside is exponential in the dimensionality of the data space. Hence, the main task of research in the field was the development of appropriate subspace search heuristics. Starting from the pioneering approaches to axis-parallel subspace clustering, two opposite basic techniques for searching subspaces were pursued, namely, top-down search[30] and bottom-up search.[31]

Top-Down Subspace Search

The rational of top-down approaches is to determine the subspace of a cluster starting from the full-dimensional space. This is usually done by determining a subset of attributes for a given set of points—potential cluster members—such that the points meet the given cluster criterion when projected onto this corresponding subspace. Obviously, the dilemma is, that for the determination of the subspace of a cluster, at least some cluster members must be identified. On the other hand, to determine cluster memberships, the subspace of each cluster must be known. To escape this circular dependency, most of the top-down approaches rely on a rather strict assumption, which has been called the *locality assumption*.[3] It is assumed that the subspace of a cluster can be derived from the local neighborhood (in the full-dimensional data space) of the cluster center or the cluster members. In other words, it is assumed that even in the full-dimensional space, the subspace of each cluster can be learned from the local neighborhood of cluster representatives or cluster members. Other top-down approaches that do not rely on the locality assumption use random sampling as a heuristic to generate a set of potential cluster members.

Bottom-Up Subspace Search

The exponential search space of all possible subspaces of a data space that needs to be traversed is equivalent to the search space of the frequent itemset problem in market basket analysis in the area of transaction databases. For example, an item set may contain items A, B, C, etc. The key idea of the APRIORI algorithm[32]

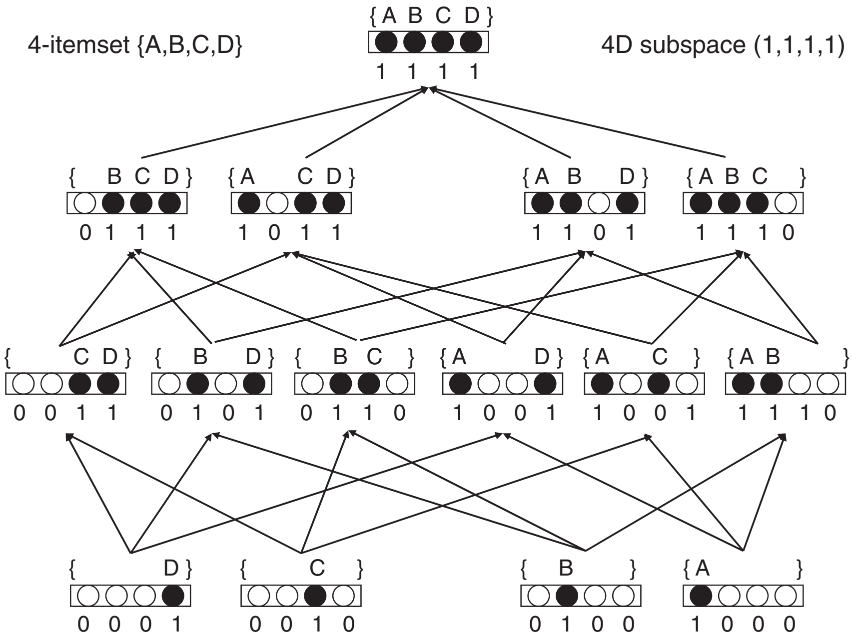


FIGURE 4 | Frequent itemset mining and subspace mining: the search space of all possible itemsets over four items {*A*, *B*, *C*, *D*} and of all possible subspaces of a four-dimensional data space where each subspace is encoded as a four-dimensional vector recording 1 for those attributes spanning the subspace and 0 for the remaining attributes.

is to start with itemsets (called 'transactions') of size 1 (containing a certain item, irrespective of other items possibly also contained in the transaction) and exclude those larger itemsets from the search that cannot be frequent anymore, given the knowledge which smaller itemsets are frequent. For example, if a 1-itemset containing A is not frequent (i.e., we count such an itemset less than a given minimum support threshold), all 2-itemsets, 3-itemsets, and so on, containing A (e.g., $\{A, B\}$, $\{A, C\}$, $\{A, B, C\}$) cannot be frequent either (otherwise itemsets containing A would have been frequent as well) and need not be tested for exceeding the minimum support threshold. Theoretically, the search space remains exponential, yet practically the search is usually substantially accelerated.

Transferring the itemset problem to subspace clustering, each attribute represents an item and each subspace cluster is a transaction of the items representing the attributes that span the corresponding subspace (cf. Figure 4). Finding itemsets with frequency 1 then relates to finding all combinations of attributes that constitute a subspace containing at least one cluster. This observation is the rational of most bottom-up subspace clustering approaches. The subspaces that contain clusters are determined starting from all one-dimensional subspaces that accommodate at least one cluster by employing a search strategy similar to frequent itemset mining algorithms. To apply any efficient frequent itemset mining algorithm, the cluster criterion must implement a downward closure property (also called monotonicity property):

If subspace S contains a cluster, then any subspace $T \subseteq S$ must also contain a cluster.

For pruning (i.e., excluding specific subspaces from consideration), the antimonotonic reverse implication can be used:

If a subspace T does not contain a cluster, then any superspace $S \supseteq T$ also cannot contain a cluster.

Let us note that there are bottom-up algorithms that do not use an APRIORI-like subspace search but, instead, apply other search heuristics.

Applying an efficient subspace search strategy addresses aspect 1 of the curse of dimensionality. Selecting a subset of attributes as relevant, corresponds to aspect 3 of the curse of dimensionality. Aspect 2 needs to be considered when adapting similarity measures to local neighborhoods and is differently important in the different approaches.

Clustering Algorithms

The two basic techniques resulted initially in two 'subspace' clustering paradigms that have been named by the pioneers in this field 'subspace clustering'[31] (in a narrower sense) and 'projected clustering'.[30] For the latter, a variant has emerged sometimes called 'soft projected clustering'.[33]

Projected Clustering

In projected clustering, which is closely related to the top-down technique of subspace search, the goal is to find a set of tuples (O_i, D_i), where O_i is the set of objects belonging to cluster i and D_i is the subset of attributes where the points O_i cluster according to a given cluster criterion. The approach introducing the task of 'projected clustering' is PROCLUS,[30] a k-medoid-like clustering algorithm. PROCLUS randomly determines a set of potential cluster centers M on a sample of points first. In the iterative cluster refinement phase, for each of the k current medoids the subspace is determined by minimizing the standard deviation of the distances of the points in the neighborhood of the medoids to the corresponding medoid along each dimension. Points are then assigned to the closest medoid considering the relevant subspace of each medoid. The clusters are refined by replacing bad medoids with new medoids from M as long as the clustering quality increases. A postprocessing step identifies noise points that are too far away from their closest medoids. The algorithm always outputs a partition of the data points into k clusters (each represented by its medoid) with corresponding subspaces and a (potentially empty) set of noise points.

Further examples of projected clustering implementing similar ideas include Refs 34–36.

Soft Projected Clustering

There is a rich literature concerned with so-called *soft* projected clustering, examples including Refs 33, 37–42. These are usually optimization approaches derived from k-means-type clustering, learning some weight vector for the different weighting of attributes. Therefore, weights w_i for attributes $i = 1, \ldots, d$ are introduced into the distance measure:

$$\|(\boldsymbol{x} - \boldsymbol{y})\|_p^{\boldsymbol{w}} = \sqrt[p]{\sum_{i=1}^{d} w_i \cdot |x_i - y_i|^p}. \qquad (3)$$

Usually, in these approaches, all weights are restricted at least to the condition

$$0 < w_i \leq 1, \qquad (4)$$

that is, to ensure the applicability of optimization techniques, no weight can become 0. In terms of subspace clustering, this means that no attribute is truly omitted and, hence, the resulting clustering resides in the full dimensional, although skewed data space. The fact that the clustering is not truly searched for in a projected space (i.e., no *hard* subspace is assigned to a specific cluster) is indicated by the term 'soft'. However, these approaches are often generally named 'subspace clustering', not reflecting the differences discussed here. Essentially, all these approaches can be seen as variants of EM clustering.[43]

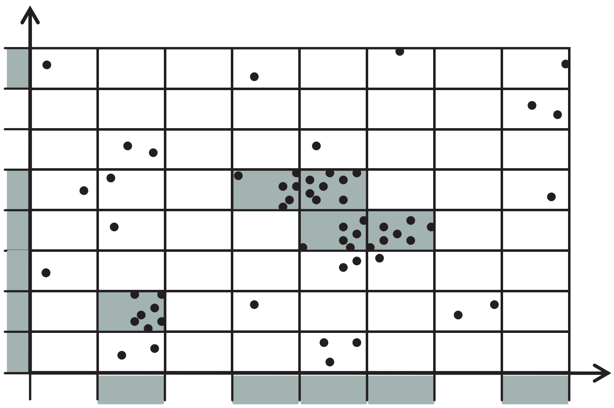


FIGURE 5 | Grid-based bottom-up subspace clustering. Dense one-dimensional and two-dimensional units (for $\tau > 3$) are highlighted. All one-dimensional projections of each dense two-dimensional unit are also dense, although not all two-dimensional combinations of dense one-dimensional units are dense.

Subspace Clustering

Subspace clustering in a narrower sense pursues the goal to find all clusters in all subspaces of the entire feature space. This goal obviously is defined to correspond to the bottom-up technique used by these approaches, based on some antimonotonic property of clusters allowing the application of the APRIORI search heuristic.

The pioneering approach for finding all clusters in all subspaces coining the term 'subspace clustering' for this specific task has been CLIQUE,[31] using a grid-based clustering notion. The data space is partitioned by an axis-parallel grid into equi-sized units of width ξ. Only units which contain at least τ points are considered as dense. A cluster is defined as a maximal set of adjacent dense units. Because dense units satisfy the downward closure property, subspace clusters can be explored rather efficiently in a bottom-up way. Starting with all one-dimensional dense units, $(k + 1)$-dimensional dense units are computed from the set of k-dimensional dense units in an APRIORI-like style. If a $(k + 1)$-dimensional unit contains a projection onto a k-dimensional unit that is not dense, then the $(k + 1)$-dimensional unit can also not be dense (see Figure 5 for an illustration). Furthermore, a heuristic that is based on the minimum description length principle is introduced to discard candidate units within less interesting subspaces, that is, subspaces that contain only a very small number of dense units. This way, the efficiency of the algorithm is enhanced but at the cost of incomplete results, i.e., possibly some true clusters are lost.

Further examples for subspace clustering in this narrower sense include Refs 44–47 as variants of CLIQUE and Ref 48 as a density-based approach also implementing the downward-closure property and an APRIORI-style search of subspaces.

The initial problem formulation of finding all clusters in all subspaces is rather questionable because the information gained by retrieving such a huge set of clusters with high redundancy is not very useful. Therefore, subsequent methods often concentrated on possibilities of concisely restricting the result set of clusters by somehow assessing and reducing the redundancy of clusters, for example, to keep only clusters of highest dimensionality. It also should be noted that the statistical significance of subspace clusters (as defined in Ref 29), is not an antimonotonic property and hence does, in general, not allow for APRIORI-like bottom-up approaches finding only *meaningful* clusters.

Hybrid Approaches

It is our impression that recently the majority of approaches does not stick to these two initial concepts any more, but pursues some hybrid approach (examples are Refs 49–55). In general, the result is neither a clear partitioning nor an enumeration of all clusters in all subspaces. Still the problem definition may remain unclear and each approach defines its own goal. A fair comparison of the true merits of different algorithms thus remains difficult.

CLUSTERING IN ARBITRARILY ORIENTED SUBSPACES

Cluster Model

A model for correlation clusters, that is, clusters residing in arbitrarily oriented subspaces, can be based on a system of linear equation describing the λ-dimensional hyperplane accommodating the points of a correlation cluster $\mathcal{C} \subset \mathbb{R}^d$. This equation system will consist of $d - \lambda$ equations for d variables, and the affinity, for example, given by the mean point $x_{\mathcal{C}} = (\bar{x}_1 \cdots \bar{x}_d)^{\mathsf{T}}$ of all cluster members:

$$\begin{gathered}
v_{1(\lambda+1)} \cdot (x_1 - \bar{x}_1) + v_{2(\lambda+1)} \cdot (x_2 - \bar{x}_2) + \cdots + v_{d(\lambda+1)} \\ \cdot (x_d - \bar{x}_d) = 0, \\
v_{1(\lambda+2)} \cdot (x_1 - \bar{x}_1) + v_{2(\lambda+2)} \cdot (x_2 - \bar{x}_2) + \cdots + v_{d(\lambda+2)} \\ \cdot (x_d - \bar{x}_d) = 0, \\
\vdots \\
v_{1d} \cdot (x_1 - \bar{x}_1) + v_{2d} \cdot (x_2 - \bar{x}_2) + \cdots + v_{dd} \\ \cdot (x_d - \bar{x}_d) = 0,
\end{gathered} \tag{5}$$

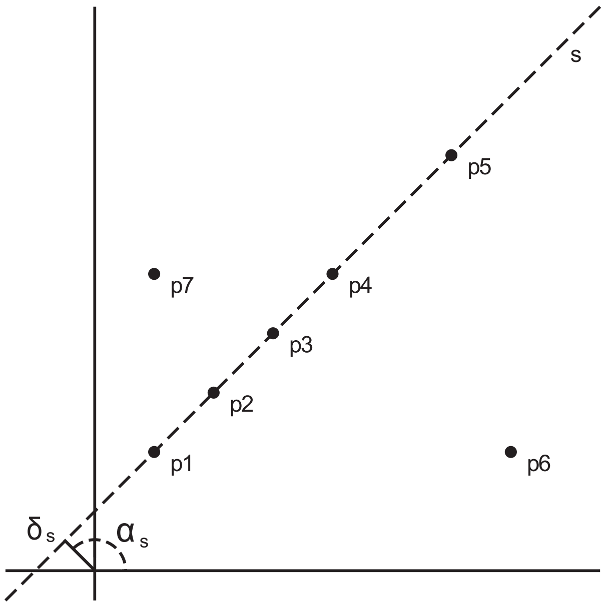
where v_{ij} is the value at row i, column j in the eigenvector matrix $V_{\mathcal{C}}$ derived by principle component analysis (PCA)[56] from the covariance matrix of $\mathcal{C}$. The first λ eigenvectors (also called *strong* eigenvectors) give the directions of high variance and span the hyperplane accommodating $\mathcal{C}$. The remaining $d - \lambda$ *weak* eigenvectors span the perpendicular subspace. The equation system is by construction at least approximately fulfilled for all points $\boldsymbol{x} \in \mathcal{C}$ and hence provides an approximate quantitative model for the cluster.[18] The degree of allowed deviation of cluster members from the hyperplane and the method of assessment differs from approach to approach. Hence, also in the area of arbitrarily oriented clustering, future research on refined and more concise models is encouraged.

Basic Techniques and Example Algorithms

Basic techniques to find arbitrarily-oriented subspaces accommodating clusters are PCA and the Hough transform. The first approach to this *generalized projected clustering* was the algorithm ORCLUS,[57] using ideas similar to the axis-parallel approach PROCLUS.[30] ORCLUS is a k-means-like approach, picking $k_c > k$ seeds at first, assigning the data objects to these seeds according to a distance function that is based on an eigensystem of the corresponding cluster assessing the distance along the *weak* eigenvectors only (i.e., the distance in the projected subspace where the cluster objects exhibit high density). The eigensystem is iteratively adapted to the current state of the updated cluster. The number k_c of clusters is reduced iteratively by merging closest pairs of clusters until the user-specified number k is reached. The closest pair of clusters is the pair with the least average distance in the projected space (spanned by the weak eigenvectors) of the eigensystem of the merged clusters. Starting with a higher k_c increases the effectiveness, but also the runtime.

Further examples of algorithms based on PCA include Refs 58–64. Some similar methods are reviewed in Ref 4. Many of these methods based on PCA use the eigensystem to adapt similarity measures in a soft way and, hence, can also be seen as variants of EM-clustering.[43]

Finding arbitrarily oriented subspace clusters by means of the Hough transform has been proposed in Refs 65 and 66. The basic idea is to map a point in the input data space (called picture space) onto a function in a so called parameter space. As depicted in Figure 6, function in the parameter space describes all possible hyperplanes of dimensionality $d - 1$ (in a d-dimensional data space) crossing a given point in



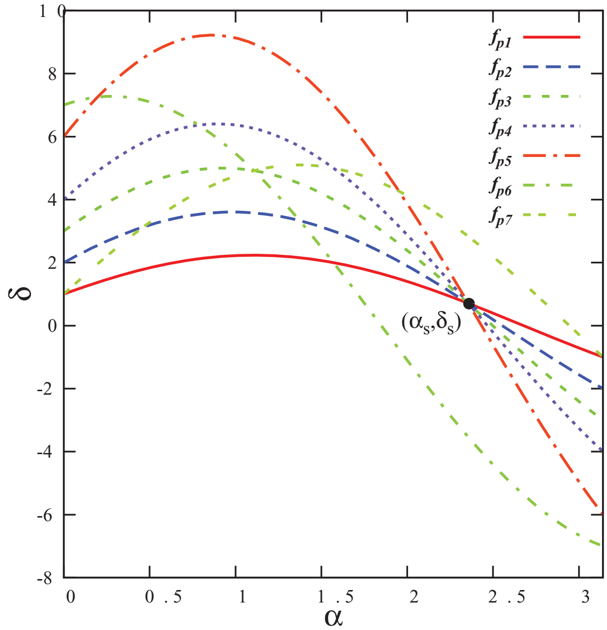


FIGURE 6 | Hough transform: points in the picture space map to functions in the parameter space, an intersection of functions in the parameter space maps to a linear manifold shared by all points mapping to these functions.

the data space. Hence, intersections of functions in the parameter space map to hyperplanes common to all points corresponding to these functions. Identifying regions of such intersections in the parameter space hence corresponds to identifying a correlation cluster in the picture space.

The Hough transform is an example of using a technique in analysis of high-dimensional data that was originally developed in the context of image analysis (hence for two-dimensional data). Other examples are the usage of image-filter techniques.[67] RANSAC[68] or other random sampling techniques have been used in Refs 69 and 70.

As mentioned, approaches to clustering in arbitrarily oriented subspaces are also known as 'correlation clustering'.[59,71] It should be noted, though, that this term is ambiguously used also for a completely different problem in machine learning where a partitioning of the data shall correlate as much as possible with a pairwise similarity function *f* learned from past data (e.g., cf. Ref 72).

CURRENT RESEARCH DIRECTIONS

Some interesting current research directions have found attention in the workshop series on 'Discovering, Summarizing and Using Multiple Clusterings' (MultiClust) at KDD 2010[73] and at ECML/PKDD 2011.[74] Intriguingly, problems known in subspace clustering meet similar problems in other areas such as ensemble clustering, alternative clustering, or multiview clustering.[75] Such problems common to all of these areas are (1) how to treat diversity of clustering solutions (are diverse clustering solutions to be unified or to be presented individually?); (2) how to effectively summarize and present diversity to a user of clustering algorithms; (3) how to treat redundancy of clusters; or (4) how to assess similarity between multiple clustering solutions. Also, relationships to frequent pattern mining, which was at the origin of subspace clustering, have been discussed from a more recent point of view on research in both fields.[76]

As we have pointed out, redundancy of subspace cluster results is a problem inherited from he bottom-up strategy of the very first approaches, borrowed from frequent pattern mining. For current research on subspace clustering, getting rid of too much redundancy is a major topic (see, e.g., Refs 77–79). Research on multiview clustering comes from the other directions,[55,80–82] seeking to allow some redundancy at least to not exclude possibly interesting concepts, although they might have a certain (partial) overlap with other concepts. A related though distinct way of tackling the problem of redundancy and distinctiveness of different clusters is to seek diverse clusterings by directly assessing a certain notion of distance between different partitions (so-called alternative clustering approaches; see, e.g., see Refs 83–90).

A question related to the redundancy issue is the appropriate density level. This is a general problem also in density-based clustering,[91] but for clustering in subspaces, the problem is aggravated. Setting a fixed density threshold for an APRIORI-style subspace search is not appropriate for all possible subspaces. Consider, for example, some CLIQUE-style grid approach: the volume of a hypercube increases exponentially with the dimensionality, hence the density decreases rapidly. As a consequence, any chosen threshold introduces a bias to identify clusters of (up to) a certain dimensionality. This observation motivates research on adaptive density thresholds (e.g., see Refs 46 and 92). When using Euclidean distance (L_2), the appropriate choice of an ε-range becomes extremely challenging as well, not only for subspace, but also for projected clustering or arbitrarily oriented clustering, due to the rather counterintuitive behavior of the volume of the hypersphere with increasing dimensions. Choosing the size of the neighborhood in terms of objects rather than in terms of a radius (i.e., using k nearest neighbors instead of an ε-range query) has been advocated as a workaround for this problem,[61] to solve at least certain aspects such as having a well-defined (nonempty) set of objects for the density estimation or spatial properties of the neighborhood.

A rather unclear problem in arbitrarily oriented clustering is the significance of arbitrarily oriented clusters. (As discussed above, for axis-parallel subspace clustering at least some first steps have been taken in Ref 29.) Also for overlapping clusters, a common result in subspace clustering but also in other domains, a proper evaluation procedure is not known. Different approaches to evaluation have different weaknesses and all neglect certain requirements. This problem is sketched and discussed in more detail in Ref 93. First steps toward improved evaluation methodology are taken in Refs 94–96.

Despite the unresolved issues in the basic techniques for real-valued feature vectors, there are also attempts to already generalize subspace clustering to more complex data. So-called three-dimensional data add a third component (e.g., time) to objects and attributes. This means, instead of a data matrix, a data tensor of third order is mined. Example approaches are Refs 97–100. As opposed to clustering of three-dimensional data, dynamic or stream data does not have complete access to the third dimension (time). There are also first attempts to address subspace clustering in dynamic or stream data.[101–103] Other challenges are given when the data are not of (purely) numeric nature but are, completely or partially, categorical,[104,105] or when the data are uncertain.[106] A recent survey discusses such problems and approaches as 'enhanced subspace clustering'.[107]

CONCLUSION

The task addressed as 'subspace clustering' attracts new approaches proposed in numerous conferences and journals every year. In this article, we distinguish different problem settings, namely, axis-parallel subspace clustering with the families of 'subspace clustering' (in a narrower sense) and '(soft) projected clustering' as well as clustering in arbitrarily oriented subspaces (also called 'correlation clustering'). Between these fields, one can find different problem settings addressed by 'biclustering' approaches. All these different families address different aspects of the so-called 'curse of dimensionality'.

Although research in the area of subspace clustering has been endured for more than 10 years, the problem is still not well defined in all aspects. It is likely that further algorithms with slightly different problem statements will be proposed. However, we would consider an important contribution to the field to carefully analyze relevant applications and to extract a meaningful problem definition that can be tackled by independently developed algorithms. After all, a clustering approach seems to be more useful if it is able to provide a model describing the clusters found. This allows us to understand the grouping, to identify underlying mechanisms quantitatively, and, eventually, to refine scientific theories or target marketing strategies.

REFERENCES

1. Parsons L, Haque E, Liu H. Subspace clustering for high dimensional data: a review. *SIGKDD Explor* 2004, 6:90–105.
2. Kröger P, Zimek A. Subspace clustering techniques. In: Liu L, Özsu MT, eds. *Encyclopedia of Database Systems*. New York, NY: Springer; 2009 2873–2875.
3. Kriegel H-P, Kröger P, Zimek A. Clustering high dimensional data: a survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Trans Knowl Discov Data* 2009, 3:1–58.
4. Vidal R. Subspace clustering. *IEEE Signal Process Mag* 2011, 28:52–68.

5. Han J, Kamber M. *Data Mining: Concepts and Techniques*. 2nd ed. San Francisco, CA: Morgan Kaufmann; 2006.
6. Gan G, Ma C, Wu J. *Data Clustering. Theory, Algorithms, and Applications*. Philadelphia, PA: Society for Industrial and Applied Mathematics; 2007.
7. Moise G, Zimek A, Kröger P, Kriegel H-P, Sander J. Subspace and projected clustering: experimental evaluation and analysis. *Knowl Inf Syst* 2009, 21:299–326.
8. Müller E, Günnemann S, Assent I, Seidl T. Evaluating clustering in subspace projections of high dimensional data. In: *Proceedings of the 35th International Conference on Very Large Data Bases (VLDB)*, Lyon, France; 2009.
9. Böhm C, Berchtold S, Keim DA. Searching in high-dimensional spaces: index structures for improving the performance of multimedia databases. *ACM Comput Surv* 2001, 33:322–373.
10. Bellman R. *Adaptive Control Processes. A Guided Tour*. Princeton, NJ: Princeton University Press; 1961.
11. Beyer K, Goldstein J, Ramakrishnan R, Shaft U. When is 'nearest neighbor' meaningful? In: *Proceedings of the 7th International Conference on Database Theory (ICDT)*, Jerusalem, Israel; 1999.
12. Hinneburg A, Aggarwal CC, Keim DA. What is the nearest neighbor in high dimensional spaces? In: *Proceedings of the 26th International Conference on Very Large Data Bases (VLDB)*, Cairo, Egypt; 2000.
13. Aggarwal CC, Hinneburg A, Keim D. On the surprising behavior of distance metrics in high dimensional space. In: *Proceedings of the 8th International Conference on Database Theory (ICDT)*, London; 2001.
14. Francois D, Wertz V, Verleysen M. The concentration of fractional distances. *IEEE Trans Knowl Data Eng* 2007, 19:873–886.
15. Bennett KP, Fayyad U, Geiger D. Density-based indexing for approximate nearest-neighbor queries. In: *Proceedings of the 5th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, San Diego, CA; 1999.
16. Houle ME, Kriegel H-P, Kröger P, Schubert E, Zimek A. Can shared-neighbor distances defeat the curse of dimensionality? In: *Proceedings of the 22nd International Conference on Scientific and Statistical Database Management (SSDBM)*, Heidelberg, Germany; 2010.
17. Bernecker T, Houle ME, Kriegel H-P, Kröger P, Renz M, Schubert E, Zimek A. Quality of similarity rankings in time series. In: *Proceedings of the 12th International Symposium on Spatial and Temporal Databases (SSTD)*, Minneapolis, MN; 2011.
18. Achtert E, Böhm C, Kriegel H-P, Kröger P, Zimek A. Deriving quantitative models for correlation clusters. In: *Proceedings of the 12th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, Philadelphia, PA; 2006.
19. Korn F, Pagel B-U, Faloutsos C. On the 'dimensionality curse' and the 'self-similarity blessing'. *IEEE Trans Knowl Data Eng* 2001, 13:96–111.
20. Fowlkes EB, Mallows CL. A method for comparing two hierarchical clusterings. *J Am Stat Assoc* 1983, 78:553–569.
21. Milligan GW. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika* 1980, 45:325–342.
22. Fowlkes EB, Gnanadesikan R, Kettenring JR. Variable selection in clustering. *J Classif* 1988, 5:205–228.
23. Gnanadesikan R, Kettenring JR, Tsao SL. Weighting and selection of variables for cluster analysis. *J Classif* 1995, 12:113–136.
24. Steinley D, Brusco MJ. Selection of variables in cluster analysis: an empirical comparison of eight procedures. *Psychometrika* 2008, 73:125–144.
25. Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003, 3:1157–1182, 2003.
26. Yang L. Distance-preserving dimensionality reduction. *WIREs Data Min Knowl Discov* 2011, 1:369–380.
27. Achtert E, Hettab A, Kriegel H-P, Schubert E, Zimek A. Spatial outlier detection: data, algorithms, visualizations. In: *Proceedings of the 12th International Symposium on Spatial and Temporal Databases (SSTD)*, Minneapolis, MN; 2011.
28. Madeira SC, Oliveira AL. Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Trans Comput Biol Bioinf* 2004, 1:24–45.
29. Moise G, Sander J. Finding non-redundant, statistically significant regions in high dimensional data: a novel approach to projected and subspace clustering. In: *Proceedings of the 14th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, Las Vegas, NV; 2008.
30. Aggarwal CC, Procopiuc CM, Wolf JL, Yu PS, Park JS. Fast algorithms for projected clustering. In: *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, Philadelphia, PA; 1999.
31. Agrawal R, Gehrke J, Gunopulos D, Raghavan P. Automatic subspace clustering of high dimensional data for data mining applications. In: *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, Seattle, WA; 1998.
32. Agrawal R, Srikant R. Fast algorithms for mining association rules. In: *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, Minneapolis, MN; 1994.
33. Jing L, Ng MK, Huang JZ. An entropy weighting *k*-means algorithm for subspace clustering of

high-dimensional sparse data. *IEEE Trans Knowl Data Eng* 2007, 19:1026–1041.
34. Woo K-G, Lee J-H, Kim M-H, Lee Y-J. FINDIT: a fast and intelligent subspace clustering algorithm using dimension voting. *Inf Softw Technol* 2004, 46:255–271.
35. Yip KY, Cheung DW, Ng MK. On discovery of extremely low-dimensional clusters using semi-supervised projected clustering. In: *Proceedings of the 21st International Conference on Data Engineering (ICDE)*, Tokyo, Japan; 2005.
36. Böhm C, Kailing K, Kriegel H-P, Kröger P. Density connected clustering with local subspace preferences. In: *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM)*, Brighton, UK; 2004.
37. Domeniconi C, Papadopoulos D, Gunopulos D, Ma S. Subspace clustering of high dimensional data. In: *Proceedings of the 4th SIAM International Conference on Data Mining (SDM)*, Lake Buena Vista, FL; 2004.
38. Friedman JH, Meulman JJ. Clustering objects on subsets of attributes. *J R Stat Soc B* 2004, 66:825–849.
39. Huang JZ, Ng MK, Rong H, Li Z. Automated variable weighting in *k*-means type clustering. *IEEE Trans Pattern Anal Mach Intell*; 2005, 27:657–668.
40. Bouveyron C, Girard S, Schmid C. High-dimensional data clustering. *Comput Stat Data Anal* 2007, 52:502–519.
41. Domeniconi C, Gunopulos D, Ma S, Yan B, M. Al Razgan, Papadopoulos D. Locally adaptive metrics for clustering high dimensional data. *Data Min Knowl Discov* 2007, 14:63–97.
42. Lu Y, Wang S, Li S, Zhou C. Particle swarm optimizer for variable weighting in clustering high-dimensional data. *Mach Learn* 2010, 82:43–70.
43. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B*, 1977, 39:1–31.
44. Cheng CH, Fu AW-C, Zhang Y. Entropy-based subspace clustering for mining numerical data. In: *Proceedings of the 5th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, San Diego, CA; 1999, 84–93.
45. Nagesh HS, Goil S, Choudhary A. Adaptive grids for clustering massive data sets. In: *Proceedings of the 1st SIAM International Conference on Data Mining (SDM)*, Chicago, IL; 2001.
46. Assent I, Krieger R, Müller E, Seidl T. DUSC: dimensionality unbiased subspace clustering. In: *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM)*, Omaha, NE; 2007.
47. Liu G, Sim K, Li J, Wong L. Efficient mining of distance-based subspace clusters. *Stat Anal Data Min* 2009, 2:427–444.
48. Kailing K, Kriegel H-P, Kröger P. Density-connected subspace clustering for high-dimensional data. In: *Proceedings of the 4th SIAM International Conference on Data Mining (SDM)*, Lake Buena Vista, FL; 2004.
49. Procopiuc CM, Jones M, Agarwal PK, Murali TM. A Monte Carlo algorithm for fast projective clustering. In: *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, Madison, WI; 2002.
50. Yip KY, Cheung DW, Ng MK. HARP: a practical projected clustering algorithm. *IEEE Trans Knowl Data Eng* 2004, 16:1387–1397.
51. Yiu ML, Mamoulis N. Iterative projected clustering by subspace mining. *IEEE Trans Knowl Data Eng* 2005, 17:176–189.
52. Kriegel H-P, Kröger P, Renz M, Wurst S. A generic framework for efficient subspace clustering of high-dimensional data. In: *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM)*, Houston, TX; 2005.
53. Achtert E, Böhm C, Kriegel H-P, Kröger P, Müller-Gorman I, Zimek A. Detection and visualization of subspace cluster hierarchies. In: *Proceedings of the 12th International Conference on Database Systems for Advanced Applications (DASFAA)*, Bangkok, Thailand; 2007.
54. Moise G, Sander J, Ester M. Robust projected clustering. *Knowl Inf Syst* 2008, 14:273–298.
55. Günnemann S, Müller E, Färber I, Seidl T. Detection of orthogonal concepts in subspaces of high dimensional data. In: *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM)*, Hong Kong, China; 2009.
56. Jolliffe IT. *Principal Component Analysis*. 2nd ed. New York, NY: Springer; 2002.
57. Aggarwal CC, Yu PS. Finding generalized projected clusters in high dimensional space. In: *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, Dallas, TX; 2000.
58. Chakrabarti K, Mehrotra S. Local dimensionality reduction: a new approach to indexing high dimensional spaces. In: *Proceedings of the 26th International Conference on Very Large Data Bases (VLDB)*, Cairo, Egypt; 2000.
59. Böhm C, Kailing K, Kröger P, Zimek A. Computing clusters of correlation connected objects. In: *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, Paris, France; 2004.
60. Li J, Huang X, Selke C, Yong J. A fast algorithm for finding correlation clusters in noise data. In: *Proceedings of the 11th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, Nanjing, China; 2007.
61. Achtert E, Böhm C, Kriegel H-P, Kröger P, Zimek A. Robust, complete, and efficient correlation clustering.

In: *Proceedings of the 7th SIAM International Conference on Data Mining (SDM)*, Minneapolis, MN; 2007.

62. Achtert E, Böhm C, Kriegel H-P, Kröger P, Zimek A. On exploring complex relationships of correlation clusters. In: *Proceedings of the 19th International Conference on Scientific and Statistical Database Management (SSDBM)*, Banff, Canada; 2007.
63. Zhang X, Pan F, Wang W. CARE: finding local linear correlations in high dimensional data. In: *Proceedings of the 24th International Conference on Data Engineering (ICDE)*, Cancun, Mexico; 2008.
64. Aziz MS, Reddy CK. A robust seedless algorithm for correlation clustering. In: *Proceedings of the 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, Hyderabad, India; 2010.
65. Achtert E, Böhm C, David J, Kröger P, Zimek A. Robust clustering in arbitrarily oriented subspaces. In: *Proceedings of the 8th SIAM International Conference on Data Mining (SDM)*, Atlanta, GA; 2008.
66. Achtert E, Böhm C, David J, Kröger P, Zimek A. Global correlation clustering based on the Hough transform. *Stat Anal Data Min* 2008, 1:111–127.
67. Cordeiro RLF, Traina AJM, Faloutsos C, Traina Jr C. Finding clusters in subspaces of very large, multidimensional datasets. In: *Proceedings of the 26th International Conference on Data Engineering (ICDE)*, Long Beach, CA; 2010.
68. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM* 1981, 24:381–395.
69. Haralick R, Harpaz R. Linear manifold clustering. In: *Proceedings of the 4th International Conference on Machine Learning and Data Mining in Pattern Recognition (MLDM)*, Leipzig, Germany; 2005.
70. Harpaz R, Haralick R. Mining subspace correlations. In: *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, Honolulu, HI; 2007.
71. Zimek A. Correlation clustering. *ACM SIGKDD Explor* 2009, 11:53–54.
72. Bansal N, Blum A, Chawla S. Correlation clustering. *Mach Learn*, 2004, 56:89–113.
73. Fern XZ, Davidson I, Dy JG. MultiClust 2010: discovering, summarizing and using multiple clusterings. *ACM SIGKDD Explor*; 2010, 12:47–49.
74. Müller E, Günnemann S, Assent I, Seidl T, editors. *Second MultiClustWorkshop: Discovering, Summarizing and Using Multiple Clusterings, Held in Conjunction with ECML PKDD 2011*, Athens, Greece; 2011.
75. Kriegel H-P, Zimek A. Subspace clustering, ensemble clustering, alternative clustering, multiview clustering: what can we learn from each other? In: *MultiClust: First International Workshop on Discovering, Summarizing and Using Multiple Clusterings Held in Conjunction with KDD 2010*, Washington, DC; 2010.
76. Vreeken J, Zimek A. When pattern met subspace cluster—a relationship story. In: *Second MultiClust Workshop: Discovering, Summarizing and Using Multiple Clusterings Held in Conjunction with ECML PKDD 2011*, Athens, Greece; 2011.
77. Assent I, Müller E, Günnemann S, Krieger R, Seidl T. Less is more: Non-redundant subspace clustering. In: *MultiClust: First International Workshop on Discovering, Summarizing and Using Multiple Clusterings Held in Conjunction with KDD 2010*, Washington, DC; 2010.
78. Günnemann S, Färber I, Müller E, Seidl T. ASCLU: alternative subspace clustering. In: *MultiClust: First International Workshop on Discovering, Summarizing and Using Multiple Clusterings Held in Conjunction with KDD 2010*, Washington, DC; 2010.
79. Müller E, Assent I, Günnemann S, Krieger R, Seidl T. Relevant subspace clustering: mining the most interesting non-redundant concepts in high dimensional data. In: *Proceedings of the 9th IEEE International Conference on Data Mining (ICDM)*, Miami, FL; 2009.
80. Bickel S, Scheffer T. Multi-view clustering. In: *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM)*, Brighton, UK; 2004.
81. Cui Y, Fern XZ, Dy JG. Non-redundant multi-view clustering via orthogonalization. In: *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM)*, Omaha, NE; 2007.
82. Jain P, Meka R, Dhillon IS. Simultaneous unsupervised learning of disparate clusterings. *Stat Anal Data Min* 2008, 1:195–210.
83. Gondek D, Hofmann T. Non-redundant data clustering. In: *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM)*, Brighton, UK; 2004.
84. Gondek D, Hofmann T. Non-redundant clustering with conditional ensembles. In: *Proceedings of the 11th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, Chicago, IL; 2005.
85. Bae E, Bailey J. COALA: a novel approach for the extraction of an alternate clustering of high quality and high dissimilarity. In: *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM)*, Hong Kong, China; 2006.
86. Davidson I, Qi Z. Finding alternative clusterings using constraints. In: *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM)*, Pisa, Italy; 2008.
87. Qi ZJ, Davidson I. A principled and flexible framework for finding alternative clusterings. In: *Proceedings of the 15th ACM International Conference on*

Knowledge Discovery and Data Mining (SIGKDD), Paris, France; 2009.

88. Davidson I, Ravi SS, Shamis L. A SAT-based framework for efficient constrained clustering. In: *Proceedings of the 10th SIAM International Conference on Data Mining (SDM)*, Columbus, OH; 2010.
89. Dang XH, Bailey J. Generation of alternative clusterings using the CAMI approach. In: *Proceedings of the 10th SIAM International Conference on Data Mining (SDM)*, Columbus, OH; 2010.
90. Phillips JM, Raman P, Venkatasubramanian S. Generating a diverse set of high-quality clusterings. In: *Second MultiClust Workshop: Discovering, Summarizing and Using Multiple Clusterings Held in Conjunction with ECML PKDD 2011*, Athens, Greece; 2011.
91. Kriegel H-P, Kröger P, Sander J, Zimek A. Density-based clustering. *WIREs Data Min Knowl Discov* 2011, 1:231–240.
92. Müller E, Assent I, Krieger R, Günnemann S, Seidl T. DensEst: density estimation for data mining in high dimensional spaces. In: *Proceedings of the 9th SIAM International Conference on Data Mining (SDM)*, Sparks, NV; 2009.
93. Färber I, Günnemann S, Kriegel H-P, Kröger P, Müller E, Schubert E, Seidl T, Zimek A. On using class-labels in evaluation of clusterings. In: *MultiClust: First International Workshop on Discovering, Summarizing and Using Multiple Clusterings Held in Conjunction with KDD 2010*, Washington, DC; 2010.
94. Patrikainen A, Meila M. Comparing subspace clusterings. *IEEE Trans Knowl Data Eng* 2006, 18:902–916.
95. Kriegel H-P, Schubert E, Zimek A. Evaluation of multiple clustering solutions. In: *Second MultiClust Workshop: Discovering, Summarizing and Using Multiple Clusterings Held in Conjunction with ECML PKDD 2011*, Athens, Greece; 2011.
96. Günnemann S, Färber I, Müller E, Assent I, Seidl T. External evaluation measures for subspace clustering. In: *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM)*, Glasgow, UK; 2011.
97. Zhao L, Zaki MJ. TRICLUSTER: an effective algorithm for mining coherent clusters in 3D microarray data. In: *Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, Baltimore, ML; 2005.
98. Sim K, Li J, Gopalkrishnan V, Liu G. Mining maximal quasi-bicliques to co-cluster stocks and financial ratios for value investment. In: *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM)*, Hong Kong, China; 2006.
99. Ji L, Tan K-L, Tung AKH. Mining frequent closed cubes in 3D datasets. In: *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB)*, Seoul, Korea; 2006.
100. Sim K, Poernomo AK, Gopalkrishnan V. Mining actionable subspace clusters in sequential data. In: *Proceedings of the 10th SIAM International Conference on Data Mining (SDM)*, Columbus, OH; 2010.
101. Aggarwal CC, Han J, Wang J, Yu PS. A framework for projected clustering of high dimensional data streams. In: *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB)*, Toronto, Canada; 2004.
102. Zhang Q, Liu J, Wang W. Incremental subspace clustering over multiple data streams. In: *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM)*, Omaha, NE; 2007.
103. Kriegel H-P, Kröger P, Ntoutsi I, Zimek A. Density based subspace clustering over dynamic data. In: *Proceedings of the 23rd International Conference on Scientific and Statistical Database Management (SSDBM)*, Portland, OR; 2011.
104. Zaki MJ, Peters M, Assent I, Seidl T. CLICKS: an effective algorithm for mining subspace clusters in categorical datasets. *Data Knowl Eng* 2007, 60:51–70.
105. Müller E, Assent I, Seidl T. HSM: heterogeneous subspace mining in high dimensional data. In: *Proceedings of the 21st International Conference on Scientific and Statistical Database Management (SSDBM)*, New Orleans, LA; 2009.
106. Günnemann S, Kremer H, Seidl T. Subspace clustering for uncertain data. In: *Proceedings of the 10th SIAM International Conference on Data Mining (SDM)*, Columbus, OH; 2010.
107. Sim K, Gopalkrishnan V, Zimek A, Cong G. A survey on enhanced subspace clustering. *Data Min Knowl Discov* 2012. doi: 10.1007/s10618-012-0258-x.

